Non - asymptotic exponential bounds for MLE deviation under minimal conditions via classical and generic chaining methods.

By Ostrovsky E., Rogover E.

Department of Mathematics and Statistics, Bar - Ilan University, 59200, Ramat Gan, Israel.
e - mail: galo@list.ru
e - mail: eugeny@soniclynx.com

Department of Mathematics and Statistics, Bar - Ilan University, 59200, Ramat Gan, Israel.
e - mail: rogovee@gmail.com

ABSTRACT

In this paper non - asymptotic exact exponential estimates are derived (under minimal conditions) for the tail of deviation of the MLE distribution in the so - called natural terms: natural function, natural distance, metric entropy, Banach spaces of random variables, contrast function, majorizing measures or, equally, generic chaining.

Key words and phrases: Risk and deviation functions, Majorizing measures, generic chaining, random variables (r.v) and fields, distance and quasi - distance, natural norm, natural metric, exponential estimations, metric entropy, maximum likelihood estimator, contrast function, integral of Hellinger, Kullback - Leibler relative entropy, partition, Young - Fenchel transform, deviation, Banach spaces of random variables, tail of distribution.

Mathematics Subject Classification (2000): primary 60G17; 62F10;
secondary 60E07; 60G70; 62F25; 62J12.

1. Introduction. Notations. Statement of problem.

Let $(\Omega, \mathcal{M}, \mathbf{P})$ be a probability space with the expectation $\mathbf{E}$, $\Omega = \{\omega\}$, $(X, \mathcal{A}, \mu)$ be a measurable space with sigma - finite non - trivial measure μ, $(\Theta, \tau) = \Theta = \{\theta\}$ be arbitrary separable local compact topological space equipped by the ordinary Borelian sigma - field, $\mathcal{F} = \{f\}$, $f = f(x, \theta)$ be a *family* of a *strictly positive* (mod μ) probabilistic densities:

$$\forall \theta \in \Theta \ \Rightarrow \int_X f(x, \theta) \ d\mu = 1,$$

$$\mu \{ \cup_{\theta \in \Theta} \{ x : \ f(x, \theta) \le 0 \} \} = 0,$$

continuous relative to the argument θ for almost all values $x; x \in X$.

We premise also the following natural condition of the identifying:

$$\forall \theta_1, \theta_2 \in \Theta, \theta_1 \ne \theta_2 \ \Rightarrow \ \mu\{x: \ f(x,\theta_1) \ne f(x,\theta_2)\} > 0.$$

Let further $\theta_0 \in \Theta$ be some fixed value of the parameter θ. We assume that $\xi = \xi(\omega)$ is a random variable (r.v) (or more generally random vector) taking the values in the space X with the density of distribution $f(x,\theta_0)$ relative the measure μ :

$$\mathbf{P}(\xi \in A) = \int_A f(x,\theta_0) \ d\mu, \ A \in \mathcal{M}. \tag{1.0}$$

The statistical sense: the r.v. ξ is the (statistical) observation (or observations) with density $f(x,\theta_0)$, where the value θ_0 is the true, but in general case unknown value of the parameter θ.

We denote as usually by $\hat{\theta}$ the Maximum Likelihood Estimation (MLE) of the parameter θ_0 based on the observation ξ :

$$\hat{\theta} = \operatorname*{argmax}_{\theta \in \Theta} f(\xi,\theta),$$

or equally

$$\hat{\theta} = \operatorname*{argmax}_{\theta \in \Theta} L(\xi,\theta) = \operatorname*{argmin}_{\theta \in \Theta}(-L(\xi,\theta)) \tag{1.1}$$

where the function

$$L = L(\xi,\theta) \stackrel{def}{=} \log\left[f(\xi,\theta)/f(\xi,\theta_0)\right]$$

is called the *contrast function*, in contradiction to the function $\theta \to f(\xi,\theta)$ or $\theta \to \log f(\xi,\theta)$, which is called ordinary *Likelihood* function.

Denote also

$$L^{(0)} = L^{(0)}(\xi,\theta) = L(\xi,\theta) \ - \ \mathbf{E}L(\xi,\theta) = L(\xi,\theta) \ - \ \mathbf{E}_0 \ L(\xi,\theta).$$

In the case if $\hat{\theta}$ is not unique, we understand as $\hat{\theta}$ any but measurable value $\hat{\theta}$ satisfying the condition (1.1).

Let now $r = r(\theta) = r(\theta,\theta_0), \ \theta \in \Theta$ be some (measurable) numerical non - negative *risk*, or *deviation* function, i.e. such that

$$r(\theta,\theta_0) \ge 0, \ r(\theta,\theta_0) = 0 \ \Leftrightarrow \theta = \theta_0,$$

not necessary to be distance, i.e. it can not satisfy the triangle inequality. We denote for arbitrary positive value v the probability of the deviation in the $r(\cdot,\cdot)$ sense of $\hat{\theta}$ from the true value θ_0 :

$$W(v) \stackrel{def}{=} \mathbf{P}(\ r(\hat{\theta},\theta_0) > v), \tag{1.2}$$

which is needed for the construction of confidence region for the unknown parameter θ_0 in the $r(\cdot,\cdot)$ sense.

Our goal of this paper is non - asymptotical estimation of the function $W = W(v)$ under minimal and natural conditions for sufficiently greatest values v, $v \ge v_0 = \text{const} > 0$, $(v >> 1)$.

Offered here estimations are some generalizations of the paper [6]. See also [1], [3], [7], [8], [22] and reference therein.

The paper is organized as follows. In the next section we introduce the needed notations and conditions. In the section 3 we describe and recall auxiliary facts about exponential bounds for tail of maximum distribution of random fields.

In the fourth section we will formulate and prove the main result of this paper. Further we consider as a particular case of smooth density function.

In the six section we consider as an application the case of *sample,* i.e. the case when the *observations* $\xi = \{\xi_i, \ i = 1, 2, \ldots, n\}$ are independent and identically distributed (i., i.d).

In the last section 7 we consider some examples in order to illustrate the precision of the obtained results.

Agreement: by the symbols $C, C_j, C(i)$ we will denote some finite positive non - essential constants.

2. Notations and conditions. Key Inequality.

It is presumed that all introduced function there exist in some domains which is described below.

$$U(v) \stackrel{def}{=} \{\theta : \ \theta \in \Theta, \ r(\theta, \theta_0) \ge v\}, \ v > 0; \tag{2.0}$$

then

$$W(v) = \mathbf{P}(\hat{\theta} \in U(v)). \tag{2.1}$$

Let $A(k), k = 1, 2, \ldots$ be some numerical strictly increasing sequence, $A(1) = 1$. For instance, $A(k) = k$ or $A(k) = k^{\Delta}, \ \Delta = \mathrm{const} > 0$ or possible $A(k) = C \ k^{\Delta}, \ k \ge k_0$. We introduce also the following measurable sets:

$$U_k = U_k(v) = \{\theta : \ \theta \in \Theta, \ r(\theta, \theta_0) \in [\ A(k) \ v, \ A(k+1) \ v \ ] \ \}.$$

We observe:

$$W(v) \le \sum_{k=1}^{\infty} W_k(v), \tag{2.2}$$

where

$$W_k(v) = \mathbf{P}\left(\hat{\theta} \in U_k(v)\right) = \mathbf{P}\left(r(\hat{\theta}, \theta) \in [\ A(k) \ v, \ A(k+1) \ v \ ]\right).$$

Introduce also the Kullback - Leibler "distance", or relative entropy, or quasi - distance between the parameters θ and θ_0 as usually

$$h(\theta) = h(\theta_0, \theta) = \mathbf{E}L(\xi, \theta) = \int_X f(x, \theta_0) \ \log[f(x, \theta)/f(x, \theta_0)] \ d\mu. \tag{2.3}$$

It is known that $h(\theta) \ge 0$ and $h(\theta) = 0 \ \Leftrightarrow \theta = \theta_0$.

We denote also

$$Y(v) = \inf_{\theta \in U(v)} h(\theta) \tag{2.4}$$

and suppose $Y(v) \in (0, \infty)$ *for all sufficiently great values* v.

Further, define the following functions (some modifications of Hellinger's integral)

$$\phi(\lambda) = \sup_{\theta \in U_1(v)} \left[\ \log \mathbf{E} \exp \left[\lambda L^{(0)}(\xi, \theta) \right] \ \right] =$$

$$\sup_{\theta \in U_1(v)} [\log \mathbf{E} \exp(\lambda L(\xi, \theta)) \ \cdot \ \exp(-\lambda h(\theta))] =$$

$$\sup_{\theta \in U_1(v)} \left[\int_X f^{\lambda}(x, \theta) \ f^{1 \ - \ \lambda}(x, \theta_0) \ d\mu \ \cdot \ \exp(-\lambda h(\theta)) \right]; \tag{2.5}$$

and

$$\chi(\lambda) = \sup_{\theta \in U(v)} \left[\ \log \mathbf{E} \exp \left[\lambda L^{(0)}(\xi, \theta) \right] \ \right] =$$

$$\sup_{\theta \in U(v)} [\log \mathbf{E} \exp(\lambda L(\xi, \theta)) \ \cdot \ \exp(-\lambda h(\theta))] =$$

$$\sup_{\theta \in U(v)} \left[\int_X f^{\lambda}(x, \theta) \ f^{1 \ - \ \lambda}(x, \theta_0) \ d\mu \ \cdot \ \exp(-\lambda h(\theta)) \right]. \tag{2.5a}$$

We suppose $\phi(\lambda) < \infty$ *or correspondingly* $\chi(\lambda) < \infty$ *for all values* λ *in some interval of a view* $(0, \lambda_0), \ \lambda_0 \in (0, \infty]$:

$$\exists \lambda_0 \in (0, \infty], \ \forall \lambda \in (0, \lambda_0] \ \Rightarrow \phi(\lambda) < \infty, \tag{2.6}$$

$$\exists \lambda_0 \in (0, \infty], \ \forall \lambda \in (0, \lambda_0] \ \Rightarrow \chi(\lambda) < \infty, \tag{2.6a}$$

Key inequality.

Note that

$$W(v) = \mathbf{P} \left(\sup_{\theta \in U(v)} L(\xi, \theta) > \sup_{\theta \notin U(v)} L(\xi, \theta) \right).$$

As long as $L(\xi, \theta_0) = 0$ and $\theta_0 \notin U$ (and $\theta_0 \notin U_1$) we conclude

$$W(v) \le \mathbf{P} \left(\sup_{\theta \in U(v)} L(\xi, \theta) > 0 \right) =$$

$$\mathbf{P} \left(\sup_{\theta \in U(v)} (L(\xi, \theta) \ - \ \mathbf{E} L(\xi, \theta)) > \inf_{\theta \in U(v)} h(\theta) \right) \le$$

$$\mathbf{P} \left(\sup_{\theta \in U(v)} L^0(\xi, \theta) > Y(v) \right) = \mathbf{P} \left(\sup_{\theta \in U(v)} [\log(f(\xi, \theta)/f(\xi, \theta_0)) \ - \ h(\theta) \ ] > Y(v) \right). \tag{2.7}$$

Therefore, we can use the known *exponentially exact* estimations of maximum random field distributions, see, for example, [1], [10], [12], [13],[14], [15], [18], [19], [20], [21] etc.

3. Auxiliary facts.

Let $(\Omega, \mathcal{M}, \mathbf{P})$ be again the probability space, $\Omega = \{\omega\}, \ T = \{t\}$ be arbitrary set, $\xi(t), \ t \in T$ be centered: $\mathbf{E}\xi(t) = 0$ separable random field (or process). For arbitrary subset $S \subset T$ we denote for the values $u \ge u_0, \ u_0 = \mathrm{const} \in (0, \infty)$

$$Q(S,u) \stackrel{def}{=} \mathbf{P}(\sup_{t \in S} \xi(t) > u); \ Q(u) := Q(T,u). \tag{3.1}$$

Our (local) goal in this section is description an exponentially exact as $u \to \infty$ estimation for the probability $Q(S,u), \ Q(u)$ in the so - called natural terms.

Definitions and some important results about $\mathbf{E} \sup_{t \in T} \xi(t)$ in general, i.e. non - Gaussian case, i.e. when the random field $\xi(t)$ may be non - Gaussian, in the terms of majoring measures see, for example, in [1], [18],[19], [20], [21]. In the so-called entropy terms this problem was considered in [5], [12], [13], [14], [15] etc.

In order to formulate our result, we need to introduce some addition notations and conditions. Let $\phi = \phi(\lambda), \lambda \in [0, \lambda_0), \ \lambda_0 = \mathrm{const} \in (0, \infty]$ be some strictly convex taking non - negative values function, such that $\phi(0) = 0$ and

$$\lambda \in [0, 0.5 \ \lambda_0) \ \Rightarrow C_1 \lambda^2 \le \phi(\lambda) \le C_2 \lambda^2; \tag{3.2}$$

$$\lim_{\lambda \to \lambda_0} \phi(\lambda)/\lambda = \infty. \tag{3.3}$$

Note that under the assumptions (2.5) or (2.5a) $\phi(\cdot) \in \Phi, \ \chi(\cdot) \in \Phi$.

We denote the set of all these function as $\Phi; \ \Phi = \{\phi(\cdot)\}$.

Further we will choose $T = U(v)$ or $T = U_k(v)$ and $\phi(\lambda)$ or correspondingly $\chi(\lambda)$ as it is defined as in (2.5) and (2.5a) .

We say that the *centered* random variable (r.v) $\xi = \xi(\omega)$ belongs to the space $B(\phi)$, if there exists some non - negative constant $\tau \ge 0$ such that

$$\forall \lambda \in [0, \lambda_0) \ \Rightarrow \mathbf{E} \exp(\lambda \xi) \le \exp[\phi(\lambda \ \tau)], \tag{3.4}$$

(the concretization of right hand side Kramer's condition).

The minimal value τ satisfying (3.4) is called a $B(\phi)$ norm of the variable ξ, write $||\xi||$ or more detail, $||\xi||B(\phi)$:

$$||\xi|| = ||\xi||B(\phi) := \inf\{\tau, \ \tau > 0, \ \forall \lambda \ \Rightarrow \mathbf{E} \exp(\lambda \xi) \le \exp(\phi(\lambda \ \tau))\}. \tag{3.5}$$

This spaces are very convenient for the investigation of the r.v. having a exponential decreasing right side tail of distribution, for instance, for investigation of the limit theorem, the exponential bounds of distribution for sums of random variables, non-asymptotical properties, problem of continuous of random fields, study of Central Limit Theorem in the Banach space etc.; see [13].

The space $B(\phi)$ with respect to the norm $|| \cdot ||B(\phi)$ and ordinary operations is a quasi - Banach space. This means by definition that:

1. $B(\phi)$ is complete relative the quasi - distance $||\xi - \eta||$;

$$\mathbf{2}.||\xi|| \ge 0; \ ||\xi|| = 0 \Leftrightarrow \xi = 0 -$$

the non - negativeness;

$$\mathbf{3}.||\xi + \eta|| \le ||\xi|| + ||\eta|| -$$

the triangle inequality;

$$\mathbf{4}.\alpha = \mathrm{const} > 0 \ \Rightarrow ||\alpha \xi|| = \alpha ||\xi|| -$$

the positive homogeneous.

The $B(\phi)$ is isomorphic to the subspace consisted on all the centered variables of quasi - Orlitz's space $(\Omega, F, \mathbf{P}), N(\cdot)$ with $N -$ right function

$$N(u) = \exp(\phi^*(u)) - 1, \ \phi^*(u) = \sup_{\lambda}(\lambda u - \phi(\lambda)).$$

The transform $\phi \to \phi^*$ is called Young - Fenchel or Legendre transform. The proof of considered assertion used the properties of saddle-point method and theorem of Fenchel - Moraux:

$$\phi^{**} = \phi.$$

Many facts about the $B(\phi)$ spaces are proved in [13], [14], p. 19 - 40:

$$\xi \in B(\phi) \Leftrightarrow \mathbf{E}\xi = 0, \ \mathbf{and} \ \exists C = \mathrm{const} > 0,$$

$$Z(\xi, x) \le \exp\left(-\phi^*(Cx)\right), x \ge 0, \tag{3.6}$$

where $Z(\xi, x)$ denotes in this article the *right hand tail* of distribution of the r.v. ξ :

$$Z(\xi, x) \stackrel{def}{=} \mathbf{P}(\xi > x), \ x \ge 0,$$

and this estimation is in general case asymptotically exact.

More exactly, if $\lambda_0 = \infty$, then the following implication holds:

$$\lim_{\lambda \to \infty} \phi^{-1}(\log \mathbf{E} \exp(\lambda \xi))/\lambda = K \in (0, \infty)$$

if and only if

$$\lim_{x \to \infty} (\phi^*)^{-1}(|\log Z(\xi, x)|)/x = 1/K.$$

Here and further $f^{-1}(\cdot)$ denotes the inverse function to the function f on the left - side half - line (C, ∞).

Let $\phi \in \Phi$. We denote

$$\phi_n(\lambda) = n\phi(\lambda/\sqrt{n}), \ \overline{\phi}(\lambda) = \sup_{n=1,2,\ldots} [n \ \phi(\lambda/\sqrt{n})] \tag{3.7}$$

and analogously

$$\chi_n(\lambda) = n\chi(\lambda/\sqrt{n}), \ \overline{\chi}(\lambda) = \sup_{n=1,2,\ldots} [n \ \chi(\lambda/\sqrt{n})]. \tag{3.7a}$$

This function obeys the following sense. If $\xi(i), \ i = 1, 2, \ldots$ be a sequence of centered, i., i.d. r.v., $\xi = \xi(1)$, belonging to the space $B(\phi)$ and having the unit norm in this space: $||\xi||B(\phi) = 1$, then we have for the normed sum

$$\eta(n) = n^{-1/2} \sum_{i=1}^{n} \xi(i) :$$

$$\mathbf{E}[\exp(\lambda \ \eta(n))] \le \exp[\phi_n(\lambda)],$$

$$\sup_{n=1,2,\ldots} \mathbf{E} \exp(\lambda \ \eta(n)) \le \exp\left[\overline{\phi}(\lambda)\right], \tag{3.8}$$

and following

$$Z(\eta(n), x) \le \exp\left[- (\phi_n)^*(x)\right],$$

the non - uniform estimation;

$$\sup_n Z(\eta(n), x) \le \exp\left[- (\overline{\phi})^*(x)\right], \tag{3.9}$$

the uniform estimation (Chernoffs estimations, see [4]).

Using the property (3.2), we can show in addition to the classical theory of the great deviations that in the "mild" zone

$$x = x(n) \in (0, C \ \sqrt{n}) \ \Rightarrow$$

$$Z(\eta(n), x) \le C_2 \ \exp\left(-C_3 \ x^2\right)$$

(the non - uniform estimation).

As an example: if in addition

$$Z(\xi(i), x) \le \exp\left(-x^q\right), \ q = \text{const} \ge 1, \ x \ge 0, \tag{3.10}$$

then for some constant $C = C(q) \in (0, \infty)$

$$\sup_{n=1,2,\ldots} Z(\eta(n), x) \le \exp\left[-C x^{\min(q,2)}\right], \tag{3.11}$$

and the last estimation is unimprovable at $x >> 1$.

Now we prove a more general assertion.

Lemma 3.1 Let $\{\eta(i)\}, \ i = 1, 2, \ldots, n$ be a sequence of i., i.d., centered r.v. such that for some $q = \text{const} > 0$ and for all positive values x

$$Z(|\eta(i)|, x) \le \exp\left(-x^q \ R(x)\right), \tag{3.12}$$

where $R(x)$ is continuous positive *slowly varying* as $x \to \infty$:

$$\forall t > 0 \ \Rightarrow \ \lim R(tx)/R(x) = 1;$$

is bounded from below in the positive semi - axis

$$\inf_{x \ge 0} R(x) > 0$$

function. For instance: $R(x) =$

$$[\log(x+3)]^r, \ R(x) = [\log(x+3)]^r \cdot [\log(\log(x+16))]^s, \ r,s = \mathrm{const}, r \ge 0.$$

Denote

$$\zeta(n) = n^{-1/2} \sum_{i=1}^{n} \eta(i).$$

We assert:

$$\sup_n Z(|\zeta(n)|, x) \le \min \left[\exp\left(-C_1(q,R) \ x^q \ R(x)\right), \ \exp\left(-C_2(q,R) \ x^2\right) \right]. \tag{3.13}$$

Notice that the lower bound, i.e. the *inverse* inequality

$$\sup_n Z(|\zeta(n)|, x) \ge \min \left[\exp\left(-C_3(q,R) \ x^q \ R(x)\right), \ \exp\left(-C_4(q,R) \ x^2\right) \right] \tag{3.14}$$

in the case when

$$Z(|\eta(i)|, x) \le \exp\left[-C_5(q,R) \ x^q \ R(x)\right]$$

and

$$Z(|\eta(i)|, x) \ge \exp\left[-C_6(q,R) \ x^q \ R(x)\right], \ 0 < C_5 \ge C_6 < \infty$$

is trivial. Namely,

$$\sup_n Z(|\zeta(n)|, x) \ge Z(|\eta(1)|, x) \ge \exp\left[-C_6(q,R) \ x^q \ R(x)\right],$$

and on the other hand

$$\sup_n Z(|\zeta(n)|, x) \ge \lim_{n \to \infty} \mathbf{P}(|\zeta(n)| > x) =$$

$$2 \int_x^\infty (2\pi)^{-1/2} \ \sigma^{-1} \ \exp\left[-y^2/(2\sigma^2)\right] \ dy \ge \exp\left[-C_7(q,R) \ x^2\right], \ x \ge C_8 > 0;$$

here we used the CLT and denoted

$$\sigma^2 = \sigma^2(q) = \mathbf{Var} \ (\eta(1)) \in (0, \infty).$$

Proof (briefly) of the Lemma 3.1.

The case $q \ge 1$ is considered in [13], chapter 1, section (1.6); therefore we must consider only the case $q \in (0,1)$.

Further, without loss of generality we can consider the case when the r.v. ξ and η are independent and symmetrical distributed with densities correspondingly

$$f_\xi(x) = f(x) = C_9 \ \exp\left(-|x|^q \ R(|x|)\right),$$

$$g_\eta(x) = g(x) = C_{10} \ \exp\left(-K \ |x|^q \ R(|x|)\right),$$

and $\tau = \xi + \eta$. Here $K = \text{const} \in (1, \infty)$ (the case when $K = 1$ may be considered analogously).

Let us assume that $x \to \infty, \ x \ge C$. We have denoting by $h(x) = h_\tau(x)$ the density of distribution of the r.v. $\tau :$

$$h(x) \sim C_{11} \int_0^x \exp\left[-K(x - y)^q R(x - y) - y^q R(y)\right] \ dy =$$

$$C_{11} x \int_0^1 \exp\left[-x^q \left[K(1 - t)^q R(x \ (1 - t)) + t^q \ R(t \ x)\right]\right] \ dt \sim$$

$$C_{11} x \int_0^1 \exp\left[-x^q \ R(x) \ S(t)\right] \ dt,$$

where

$$S(t) = K(1 - t)^q + t^q.$$

The function $t \to S(t), \ t \in [0,1]$ achieves the minimal value K at the (critical) point $t = 0$ and as $t \to 0+$

$$S(t) = K + t^q + 0(t).$$

Note that in the case $K = 1$ there are two critical points: $t = 0$ and $t = 1$.

Further, we use the classical saddle - point method: at $x \to \infty, x > 1$ we have:

$$h(x) \sim C_{11} x \int_0^\infty \exp\left[-x^q \ R(x) \ (K + t^q)\right] \ dt =$$

$$C_{12}(q, R) \ x \ \exp\left[-K \ x^q \ R(x)\right] \ \left(x^q \ R(x)\right)^{-1/q} \le$$

$$C_{13}(q, R) \ \exp\left[-K \ x^q \ R(x)\right].$$

This completes the proof of the lemma 3.1.

The function $\phi(\cdot)$ may be introduced constructive, i.e. *only by means of the values of the considered random field* $\{\xi(t), \ t \in T\}$ by the formula

$$\phi(\lambda) = \phi_0(\lambda) \stackrel{def}{=} \log \sup_{t \in T} \mathbf{E} \exp(\lambda \xi(t)), \tag{3.15}$$

if obviously the family of the centered r.v. $\{\xi(t), \ t \in T\}$ satisfies the *uniform* Kramers condition:

$$\exists C \in (0, \infty), \ \sup_{t \in T} Z(\xi(t), \ x) \le \exp(-C \ x), \ x \ge 0. \tag{3.16}$$

In this case, i.e. in the case the choice the function $\phi(\cdot)$ by the formula (3.15), we will call the function $\phi(\lambda) = \phi_0(\lambda)$ a *natural* function.

Note that if for some $C = \mathrm{const} \in (0,\infty)$

$$Q(T,u) \le \exp\left(-\phi^*(Cu)\right),$$

then the condition (2.6) is satisfied (the necessity of the condition (2.6)).

M.Talagrand [18], [19], [20], [21], W.Bednorz [2], X. Fernique [5] etc. write instead our function $\exp\left(-\phi^*(x)\right)$ the function $1/\Psi(x)$, where $\Psi(\cdot)$ is some Youngs function and used as a rule a function $\Psi(x) = \exp(x^2/2)$ (the so - called subgaussian case).

Without loss of generality we can and will suppose

$$\sup_{t\in T}[\ ||\xi(t)\ ||B(\phi)] = 1,$$

(this condition is satisfied automatically in the case of natural choosing of the function $\phi:\ \ \phi(\lambda) = \phi_0(\lambda)$) and that the metric space (T,d) relatively the so-called *natural* distances (more exactly, semi - distances)

$$d_\phi(t,s) = d(t,s) \stackrel{def}{=} ||\xi(t)\ -\ \xi(s)||B(\phi), \tag{3.17}$$

and analogously (see the definition of a function $\chi(\cdot)$ further)

$$d_\chi(t,s) \stackrel{def}{=} ||\xi(t)\ -\ \xi(s)||B(\chi) \tag{3.17a}$$

is complete.

Recall that the semi - distance $\rho = \rho(t,s),\ s,t \in T$, for instance, $d = d_\phi(t,s),\ s,t \in T$ is, by definition, a non - negative symmetrical numerical function, $\rho(t,t) = 0,\ t \in T$, satisfying the triangle inequality, but the equality $\rho(t,s) = 0$ does not means (in general case) that $s = t$.

For example, if the random field $\xi(t)$ is centered and normed:

$$\sup_{t\in T} \mathbf{Var}\ [\xi(t)] = 1$$

Gaussian field with a covariation function $D(t,s) = \mathbf{E}\xi(t)\ \xi(s)$, then $\phi_0(\lambda) = 0.5\ \lambda^2,\ \lambda \in R$, and $d(t,s) =$

$$||\xi(t) - \xi(s)||B(\phi_0) = \sqrt{\mathbf{Var}[\xi(\mathbf{t}) - \xi(\mathbf{s})]} = \sqrt{D(t,t)\ -\ 2D(t,s) + D(s,s)}.$$

Let (T,ρ) be a compact metrical space. We us introduce as usually for any subset $V,\ V \subset T$ the so-called *entropy* $H(V,\rho,\epsilon) = H(\rho,\epsilon)$ as a logarithm of a minimal quantity $N(V,\rho,\epsilon) = N(V,\epsilon) = N(\rho,\epsilon)$ of a balls in the distance $\rho(\cdot,\cdot)$ $S(V,t,\epsilon),\ t \in V$:

$$S(V,t,\epsilon) \stackrel{def}{=} \{s, s \in V,\ \rho(s,t) \le \epsilon\},$$

which cover the set V :

$$N = \min\{M : \exists \{t_i\}, i = 1,2,,M,\ t_i \in V,\ V \subset \cup_{i=1}^M S(V,t_i,\epsilon)\},$$

and we denote also

$$H(V,\rho,\epsilon) = \log N;\ S(t_0,\epsilon) \stackrel{def}{=} S(V,t_0,\epsilon)$$

$$H(\rho, \epsilon) \stackrel{def}{=} H(T, \rho, \epsilon). \tag{3.18}$$

It follows from Hausdorf's theorem conversely that $\forall \epsilon > 0 \ \Rightarrow H(V, \rho, \epsilon) < \infty$ iff the metric space (V, ρ) is precompact set, i.e. is the bounded set with compact closure.

We quote now some results from [13], [14], [15] about the non - asymptotic exponential estimations for $Q(u) = Q(T, u)$ as $u >> 1$. Define for any value $\delta \in (0, 1)$ and arbitrary subset V of the space $\Theta : \ V \subset \Theta$ and some semi - distance $\rho(\cdot, \cdot)$ on the set T the following function:

$$G(V, \rho, \delta) = G(\rho, \delta) = \sum_{m=1}^{\infty} \delta^{m \ - \ 1} \ \cdot \ H(V, \rho, \delta^m) \ \cdot \ (1 \ - \ \delta). \tag{3.19}$$

We define formally $G(\delta) = +\infty$ for the values $\delta > \delta_0$.

In the case when $V = U(v)$ and $\rho(t, s) = d_{\phi}(t, s)$, i.e. when ρ is the natural semi - distance, we will write for brevity $G(\delta) = G(U(v), d_{\phi}, \delta)$.

If

$$\exists \delta_0 \in (0, 1), \ \forall \delta \in (0, \delta_0) \ \Rightarrow G(\delta) < \infty, \tag{3.20}$$

then

$$Q(T, u) \le V(T, \delta, u), \ V(T, \delta, u) \stackrel{def}{=} \exp\left(G(\delta) \ - \ \phi^*(u(1 \ - \ \delta)) \right), \tag{3.21}$$

or equally

$$Q(T, u) \le \inf_{\delta \in (0,1)} V(T, \delta, u). \tag{3.22}$$

If for example

$$\forall \delta \in (0, 1/e] \ \Rightarrow G(T, d_{\phi}, \delta) \le H_0 + \kappa |\log \delta|, \ H_0, \kappa = \text{const} < \infty,$$

then we get denoting

$$\pi(u) = u \ \phi^{*/}(u)$$

for the values u for which $\pi(u) \ge 2\kappa$:

$$Q(T, u) \le \exp(H_0) \ C^{\kappa} \ \kappa^{-\kappa} \ (\pi(u))^{\kappa} \ \exp\left(-\phi^*(u) \right), \tag{3.23}$$

and the last estimation (3.23) is exact in the main (exponential) term $\exp((-\phi^*(u))$.

More exactly, in many practical cases the following inequality holds:

$$\forall \epsilon \in (0, 3/4) \ \exists K > 0, \ \forall u > K \ \Rightarrow \pi(u) < \exp(\ - \ \phi^*(\epsilon \ u)); \tag{3.24}$$

and we conclude hence for $u > K = K(\epsilon)$ by virtue of convexity of a function $\phi^*(x)$:

$$Q(T, u) \le \ C_1(\kappa, \phi(\cdot)) \ \exp\left(-\phi^*((1 \ - \epsilon) \ u) \right),$$

and conversely there exists a r.v. ζ with unit norm in the space $B(\phi) : \ \zeta \in B(\phi), \ ||\zeta|| = 1,$ for which

$$u > K \ \Rightarrow \ Z(\zeta, u) \ge \ C_2(\phi) \ \exp\left(-\phi^*((1 \ + \epsilon) \ u) \right)$$

The value κ is called the *metric dimension* of the set T relative the distance $d = d_\phi(\cdot, \cdot)$.
Note that if

$$T = \cup_{m=1}^\infty T(m)$$

is some measurable partition $R = \{T(m)\}$ of the parametrical set T, then

$$Q(T, u) \le \sum_{m=1}^\infty Q(T_m, u)$$

and hence

$$Q(T, u) \le \inf_{R=\{T(m)\}} \sum_{m=1}^\infty Q(T_m, u).$$

Estimating the right side term by means of the inequality (3.19), we get:

$$Q(T, u) \le \inf_{R=\{T(m)\}} \left[\sum_{m=1}^\infty \inf_{\delta(m) \in (0,1)} \sum_{m=1}^\infty V(T(m), \delta(m), u) \right]. \tag{3.24}$$

The last assertion is some simplification of the *Majorizing Measures, or Generic Chaining Method* (see [5], [18] - [21], [2], [11] etc).

Further we will use as a rule the partition R of the set $U(v)$ of a view

$$R = \cup_{k=1}^\infty \{\theta : \theta \in U(v),\ r(\theta, \theta_0) \in [\ A(k)\ v,\ A(k+1)\ v\ ]\ \}. \tag{3.25}$$

4. Main results.

A. Compact parametrical set.

The compactness means by definition that the function $\theta \to r(\theta, \theta_0), \theta \in \Theta$ is bounded. Since as a rule the parametric set Θ is a closed subset in Euclidean finite - dimensional space and $r(\cdot, \cdot)$ is ordinary distance, this definition coincides with usually definition of the compact sets.

Note that in this case only finite numbers of the sets $\{A(k)\}$ are non - empty. We can suppose in this subsection for simplicity $U_1(v) = U(v)$ and therefore $\phi(\lambda) = \chi(\lambda)$.

Let the function $\phi = \phi(\lambda)$ be defined as in (2.5a) or equally (in the considered case) as in (2.5). Recall that

$$\sup_{\theta \in U(v)} ||L^0(\xi, \theta)|| B(\chi) = 1.$$

Introduce the so - called *natural* semi - distance on the set $U(v)$ as follows:

$$d = d_\chi = d(\theta_1, \theta_2) = ||L^0(\xi, \theta_1) \ - \ L^0(\xi, \theta_2)|| B(\chi) =$$

$$|| \log[f(\xi, \theta_1)/f(\xi, \theta_2)] \ - \ [h(\theta_1) \ - \ h(\theta_2)] || B(\chi). \tag{4.1}$$

It follows immediately from (3.18) (or equally from (3.19)) the following result.

Theorem 4.1.a.. If there exists $\delta_0 = \text{const} \in (0, 1)$ such that $\forall \delta \in (0, 1) \ \Rightarrow$

$$G(U(v), d_\chi, \delta) := \sum_{m=1}^{\infty} \delta^{m - 1} H(U(v), d_\chi, \delta^m) (1 - \delta) < \infty, \tag{4.2),}$$

then $\forall \delta \in (0, \delta_0]$

$$W(v) \le \exp \left[G(U(v), d_\chi, \delta) - \chi^*((1 - \delta) Y(v)) \right]. \tag{4.3}$$

Let us offer the more convenient for application form. Define for $\tilde{U} \subset U(v)$, arbitrary function $\nu \in \Phi$, and any semi - distance $\rho = \rho(\theta_1, \theta_2)$ on the set $\tilde{U}$ the following function (if it is finite)

$$\Psi_\nu(\tilde{U}, \rho, y) \stackrel{def}{=} \inf_{\delta \in (0,1)} \exp \left[G(\tilde{U}, \rho, \delta) - \nu^*((1 - \delta) y) \right]. \tag{4.4}$$

Theorem 4.1. Under the conditions of the theorem (4.1.a) the following estimate is true:

$$W(v) \le \Psi_\chi(U(v), d_\chi, Y(v)). \tag{4.5}$$

B. Non - compact set.

In this case we need to use the main idea of the so - called generic chaining , or majorizing measure method (3.22), (see [5], [18] - [21], [2], [11] etc), which used in particular the partition $U(v) = \cup_k U_k(v)$.

Let us denote for the partition $R = \{U_k(v)\}, U(v) = \cup_k U_k(v)$

$$\tau_k = \tau_k(v) = \sup_{\theta \in U_k(v)} ||L^0(\xi, \theta)|| B(\phi),$$

$$Y_k(v) = \inf_{\theta \in U_k(v)} h(\theta),$$

and introduce the following distance d_k on the set $U_k = U_k(v)$:

$$d_k \left(\theta_1^{(k)}, \theta_2^{(k)} \right) = ||L^0(\xi, \theta_1^{(k)}) - L^0(\xi, \theta_2^{(k)})|| B(\phi), \theta_1^{(k)}, \ \theta_2^{(k)} \in U_k(v).$$

Theorem 4.2. We have for arbitrary partition R

$$W(v) \le \sum_{k=1}^{\infty} \Psi_\phi \left(U_k(v), \frac{d_k}{\tau_k(v)}, \frac{Y_k(v)}{\tau_k(v)} \right). \tag{4.6}$$

Notice that

$$H(V, \rho/K, \epsilon) = H(V, \rho, K \cdot \epsilon), \ K = \text{const} > 0. \tag{4.7).}$$

Proof of the Theorem 4.2. We use the inequality (2.2): $W(v) \le \sum_k W_k(v)$. Let us estimate each summand $W_k(v)$:

$$W_k(v) = \mathbf{P} \left(\sup_{\theta \in U_k(v)} L(\xi, \theta) > 0 \right) =$$

$$\mathbf{P}\left(\sup_{\theta \in U_k(v)} [L^0(\xi,\theta) \ - \ h(\theta)] > 0\right) \le$$

$$\mathbf{P}\left(\sup_{\theta \in U_k(v)} L^0(\xi,\theta) > Y_k(v)\right) = \mathbf{P}\left(\sup_{\theta \in U_k(v)} \frac{L^0(\xi,\theta)}{\tau_k(v)} > \frac{Y_k(v)}{\tau_k(v)}\right). \tag{4.8}$$

The random field

$$\xi_k(\theta) = \frac{L^0(\xi,\theta)}{\tau_k(v)}, \ \theta \in U_k(v)$$

is normed in the $B(\phi)$ sense:

$$\sup_{\theta \in U_k(v)} ||\xi_k(\theta)||B(\phi) = 1.$$

Further,

$$||\xi_k(\theta_1^{(k)}) \ - \ \xi_k(\theta_2^{(k)})||B(\phi) = d_k\left(\theta_1^{(k)}, \theta_2^{(k)}\right).$$

Using the inequality (3.22) for the probability $W_k(v)$ and summing over k, we arrive to the estimation (4.6).

5. The regular, or smooth case.

A. Non - formal introduction. Restrictions. Conditions.

In this section we consider the case when the set Θ is closed (may be unbounded) convex nonempty subset of the Euclidean space $R^m, \ m = 1, 2, \ldots,$ the density $f(x,\theta)$ is twice differentiable function on the variable (variables) θ.

We choose as the deviation function *hereafter* $r(\theta,\theta_0)$ the ordinary Euclidean distance

$$r(\theta_1,\theta_2) = \sqrt{(\theta_1 \ - \ \theta_2, \theta_1 \ - \ \theta_2)} \stackrel{def}{=} |\theta_1 \ - \theta_2|.$$

The function ϕ is in this section the natural, i.e. $\phi(\lambda) = \phi_0(\lambda)$.

We have formally as $\theta \to \theta_0$, denoting $\nabla \ f = \operatorname{grad}_\theta \ f = \partial f / \partial \theta :$

$$h(\theta) \sim \int_X f(x,\theta_0) \ \times$$

$$\log\left(\frac{f(x,\theta_0) + \nabla f(x,\theta_0)(\theta \ - \ \theta_0) + 0.5 \nabla^2 \ f(x,\theta_0)(\theta - \theta_0, \theta \ - \ \theta_0)}{f(x,\theta_0)}\right) \ \mu(dx) \asymp$$

$$C \ r(\theta,\theta_0)^2 = C \ |\theta \ - \ \theta_0|^2, \ C = C(f(\cdot,\cdot),\theta_0).$$

It is reasonable to assume that

$$h(\theta) \asymp C \ |\theta \ - \ \theta_0|^2. \tag{5.1}$$

B. Main result of this section.

Theorem 5.1.

We impose on the our statistical structure the following conditions.

A. Let the function $\phi(\lambda) = \phi_0(\lambda)$ satisfied the condition (2.6) on the set $T = U(1)$.

B. Assume that the condition (5.1) is satisfied.

C. Suppose there exists a constant $C > 1$ such that for each constant $K > 1$ the following inequality holds:

$$\sup_{\theta: v \le |\theta \ - \ \theta_0| \le K \ v} \frac{||L^0(\xi, \theta)||}{|\theta \ - \ \theta_0|} \le C \cdot K; \tag{5.2}$$

Then there exists a constant $C = C(f(\cdot, \cdot), m, \theta_0) \in (0, \infty)$ *such that for all the values* $v \ge 1$

$$W(v) \le \exp\left(-\phi^*(C \cdot v)\right). \tag{5.3}$$

Proof.

1. We intend to use the result of the theorem 4.2. First of all we choose the partition R of a view: $R = \cup_k [\ A(k)\ v, \ A(k+1)\ v\ ]$, where $A(k) = k, \ k = 1, 2, \ldots.$

2. From the conditions **B**, or equally the condition (5.1) and the condition **C** follows that:

$$\tau_k(v) \le C_2 \ (k+1) \ v \tag{5.4}$$

and

$$Y_k \stackrel{def}{=} Y_k(v) \ge C_3 A_k^2 v^2. \tag{5.5}$$

3. Since the function $\phi(\cdot) = \phi_0(\cdot)$ satisfies the condition **A** , we can estimate the natural distance d_k as follows:

$$d_k\left(\theta_1^{(k)}, \theta_2^{(k)}\right)/\tau_k = ||L^0(\xi, \theta_1^{(k)}) \ - \ L^0(\xi, \theta_2^{(k)})||B(\phi)/\tau_k \le$$

$$C_4 \ |\theta_1^{(k)} \ - \ \theta_2^{(k)}|. \tag{5.6}$$

Since the layer U_k is bounded in the Euclidean metric, we conclude from (5.6) that

$$H(U_k(v), d_k/\tau_k, \delta) \le C_6(L, m) + m \ |\log \delta|. \tag{5.7}$$

On the other words, in the considered regular case $\kappa = m.$

Therefore, all the conditions of theorem 4.2 are satisfied, and we obtain from the inequality (4.6): $W(v) \le W_0(v)$, where

$$W_0(v) \stackrel{def}{=} \sum_{k=1}^{\infty} \exp\left(-\phi^*(k^2 \ v^2/(C_7 \ (k+1) \ v))\right) \le$$

$$\sum_{k=1}^{\infty} \exp\left(-\phi^*(C_8 \ k \ v)\right) \ \le \ \exp\left(-\phi^*(C_9 \ v)\right) \tag{5.8}$$

as long as $v \ge 1$.

This completes the proof of theorem 5.1.

Corollary 5.1. The conclusion of the theorem (5.1), i.e. the inequality (5.3) may be rewritten as follows. For all the values $v \ge 0$

$$W(v) \le \min\left[1, W_0(v)\right]. \tag{5.9}$$

Note that $W_0(0) = +\infty$.

Corollary 5.2.
We obtain using the asymptotical behavior of the function $\phi = \phi(\lambda), \lambda \to 0+$ in the *bounded interval* of the variable $v: \ v \in [1, C_1], \ C_1 = \mathrm{const} > 1$

$$W(v) \le \exp\left(-C \ v^2\right). \tag{5.10}$$

Notice that under some additional conditions, see [7], chapter 3, section 3, at $v \le 1$ the following inequality holds:

$$W(v) \le \exp\left(-C_2 \ v^2\right).$$

Therefore, we get under these conditions at $v \le C_3 = \mathrm{const} > 0$

$$W(v) \le \exp\left(-C_4 \ v^2\right). \tag{5.11}$$

Remark 5.1 We conclude in the smooth case, taking the union of inequalities 5.8 and 5.11 and taking into account the behavior of the function $\phi^*(\lambda)$ as $\lambda \to 0+$: as in the case of the of the function $\phi(\lambda)$:

$$\phi^*(\lambda) \sim \lambda^2, \ \lambda \in [0, C],$$

that

$$W(v) \le \exp\left[-\phi^*(C \ v)\right]. \tag{5.12}$$

We obtained the main result of this report.

6. The case of sample.

In this section we consider the case when $\xi = \vec{\xi} = \{\xi(i)\}, i = 1, 2, \ldots, n$ are i., i.d. r.v. with the (one - dimensional) density $f(x, \theta)$, *satisfying all the condition of the sections 1 and 5*, (the smooth case.)

We keep also all notations for the function $f(\cdot, \cdot)$, for instance the notions $h(\theta), Y(v), \phi(\cdot), \overline{\phi}, \ R$ etc.

We will investigate in this section the *non - uniform* probability *under natural norming* $\sqrt{n}$:

$$W_n(v) = \mathbf{P}(\sqrt{n} \ r(\hat{\theta_n}, \theta_0) > v), \tag{6.1}$$

where $\hat{\theta}_n = \hat{\theta}$ is the MLE estimation of the unknown parameter θ_0 on the basis the sample $\xi = \vec{\xi}$:

$$\hat{\theta} = \hat{\theta_n} = \operatorname*{argmax}_{\theta \in \Theta} \prod_{i=1}^{n} f(\xi(i), \theta)$$

or equally

$$\hat{\theta} = \operatorname*{argmax}_{\theta \in \Theta} L(\xi, \theta) = \operatorname*{argmin}_{\theta \in \Theta} (-L(\xi, \theta)) \tag{6.2}$$

where the contrast function $L(\cdot, \cdot)$ may be written here as

$$L = L^{(n)} = L^{(n)}(\xi, \theta) = \sum_{i=1}^{n} \log \left[f(\xi(i), \theta) / f(\xi(i), \theta_0) \right] \tag{6.3}$$

and correspondingly

$$L_0 = L_0^{(n)} = L^{(n)}(\xi, \theta) = \sum_{i=1}^{n} \log \left[f(\xi(i), \theta) / f(\xi(i), \theta_0) \ - \ h(\theta) \right] \tag{6.3.a}$$

and find also the upper estimation for the *uniform* probability

$$\overline{W}(v) = \sup_n W_n(v). \tag{6.4}$$

Theorem 6.1 Under the formulated conditions the following estimations are true:

$$W_n(v) \le \exp\left(-\phi_n^*(C_1 \ v)\right), \tag{6.5}$$

$$\overline{W}(v) \le \exp\left(-\overline{\phi}^*(C_1 \ v)\right). \tag{6.6}$$

Proof. Let us denote for brevity

$$\eta(i, \theta) = \eta(i) = \log \left[f(\xi(i), \theta) / f(\xi(i), \theta_0) \right],$$

$$\eta^o(i) = \eta^o(i, \theta) = \eta(i) \ - \ \mathbf{E}\eta(i) = \eta(i) \ - \ h(\theta).$$

We have using the key inequality for the sample of a volume n :

$$W_n(v) \le \mathbf{P}\left(\sup_{\theta \in U(v/\sqrt{n})} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta(i, \theta) > 0 \right) \le$$

$$\mathbf{P}\left(\sup_{\theta \in U(v/\sqrt{n})} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\eta^o(i, \theta) \ - \ h(\theta) \right] > 0 \right) \le$$

$$\mathbf{P}\left(\sup_{\theta \in U(v/\sqrt{n})} \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{n} \eta^o(i, \theta) \right] > \sqrt{n} \ Y(v/\sqrt{n}) \right) =$$

$$\mathbf{P}\left(\sup_{\theta \in U(v/\sqrt{n})} \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{n} \eta^o(i, \theta) \right] / \tau(v/\sqrt{n}) > \sqrt{n} \ Y(v/\sqrt{n}) / \tau(v/\sqrt{n}) \right). \tag{6.7}$$

As long as

$$Y(v) \ge C_1 \ v^2, \ \tau(v) \stackrel{def}{=} \tau_1(v) \le C_2 \ v, \ v > 0$$

we can use for the estimation of the distribution of the r.v.

$$\zeta_n(\theta) \stackrel{def}{=} \frac{1}{\sqrt{n}} \left[\sum_{i=1}^n \eta^o(i,\theta) \right] / \tau(v/\sqrt{n})$$

and the difference

$$\zeta_n(\theta_1) \ - \ \zeta_n(\theta_2) = \frac{1}{\sqrt{n}} \left[\sum_{i=1}^n (\eta^o(i,\theta_1) \ - \ \eta^o(i,\theta_2)) \right] / \tau(v/\sqrt{n}) \tag{6.8}$$

the definition of the function $\phi^*(\cdot)$ and its properties; another approach in the many general cases, i.e. when the function $\phi(\cdot)$ does not exists, may be investigated by means of the Lemma 3.1.

Using the estimation (5.12), we affirm

$$W_n(v) \le \exp\left(-\phi_n^* \left[C \ v\right]\right). \tag{6.9}$$

The second assertion of the theorem 6.1 follows immediately by passing to $\sup_n$.

Remark 6.1 From the assertion of the theorem 6.1 it may be obtained the estimations from *integral* measures of deviation. For instance, if we choose the loss function $l(\cdot)$ of a kind

$$l = \sqrt{n} \ r(\hat{\theta_n}, \theta),$$

then

$$\exists C \in (0,\infty), \ \sup_n ||\sqrt{n} \ r(\hat{\theta_n},\theta)||B(\overline{\phi}) = C(f(\cdot,\cdot)) < \infty. \tag{6.10}$$

As a corollary: for all values $p = \text{const} \in [1,\infty)$

$$\sup_n |\sqrt{n} \ r(\hat{\theta_n},\theta)|_p \le C_1 \ p/\overline{\phi}^{-1}(p) < \infty, \tag{6.11}$$

where we used the ordinary notation: for arbitrary r.v. ζ

$$|\zeta|_p \stackrel{def}{=} [\mathbf{E}|\zeta|^p]^{1/p}.$$

7. Some examples.

Example 7.1. Spherical unimodal distributions.

We consider the following first example (and other examples) in order to illustrate the precision of the theorems 4.1, 4.2 and 5.1.

Let $q = \text{const} \ge 2, \ X = R^m, \ \mu$ be an usually Lebesgue measure, $x \in R^m \ \Rightarrow \ |x| = (x,x)^{1/2}; \ R(y), y \in [0,\infty)$ be twice continuous differentiable strictly positive:

$$\inf_{y \in [0,1]} R(y) > 0, \ \inf_{y \in [1,\infty)} y^q R(y) > 0,$$

slowly varying as $y \to \infty$ functions such that the function $y \to \ y^q R(y), \ y \ge 0$ is strictly monotonically increasing

Let us introduce the following density function

$$f_0(x) = C(q,m,R)\ \exp\left(-|x|^q\ R(|x|)\right),$$

where $C(q,m,R)$ is a norming constant:

$$\int_{R^m} f_0(|x|)\ dx = 1.$$

We take as a parametric set $\Theta = X = R^m$; choose $\theta_0 := 0$, and define the *family* of a densities of a view (shift family):

$$f(x,\theta) = f_0(|x\ -\ \theta|),\ \theta \in \Theta = R^m. \tag{7.1}$$

Recall that the observation (observations) ξ has (have) the density of distribution $f_0(|x|)$.

It follows from the *unimodality* of the density function that the MLE of the parameter θ coincides with the observation ξ :

$$\hat{\theta} = \xi. \tag{7.2}$$

A. Upper bound.

It follows after some computations on the basis of the theorem 5.1 that (using the classical results from the theory of slowly, or regular varying functions functions) (see [16], pp. 41 - 53) that for the function $\phi(\lambda) = |\lambda|^q\ R(|\lambda|)$ the Young - Fenchel transform has a following asymptotic: as $\lambda \to\ \infty\ \Rightarrow$

$$(|\lambda|^q\ R(|\lambda|))^* \sim C\ |\lambda|^p / R\left(|\lambda|^{p\ -\ 1}\right),$$

where as usually $1/p + 1/q = 1$. As long as $q \ge 2$, we conclude that $p \in (1,2]$.

We obtain on the basis of theorem 5.1: $v \ge 1\ \Rightarrow$

$$W(v) \le \exp\left[-C_3\ v^p\ / R\left(v^{p\ -\ 1}\right]\right). \tag{7.3}$$

B. Low bound.

We get using the explicit representation (7.2) and passing to the polar coordinates:

$$W(v) = \mathbf{P}(|\xi| > v) = C(q,m,R) \int_{x:|x|>v} \exp\left(-|x|^q\ R(|x|)\right)\ dx =$$

$$C_9(q,m,R) \int_v^\infty y^{m\ -\ 1}\ \exp\left(-y^q\ R(y)\right)\ dy \ge$$

$$C_{10}\ \exp\left(-v^q\ R(v)/C_{11}\right). \tag{7.4}$$

Notice that the upper (7.3) and low bounds (7.4) exponential coincides if for instance $p = q = 2$ and $R = \text{const}$ (the Gaussian case).

Analogously may be considered a more general case of the classical MLE estimations.

Example 7.2. Smooth sample.

We suppose here that all the conditions of the theorem 6.1. are satisfied.

It follows from the formula (6.8) that

$$W_n(v) \le \exp\left(-n\ \phi^*(C_1\ v/\sqrt{n})\right). \tag{7.5}$$

Assume that the variable v belongs to the following zone: for some nonrandom positive constant $C < \infty$

$$v \le C_2\ \sqrt{n} \tag{7.6}$$

(a big zone of great deviations). Substituting into (7.5) and taking into account the behavior of the function $\phi^* = \phi^*(\lambda)$ we obtain in the considered zone the estimation:

$$W_n(v) \le \exp\left(-C_3\ v^2\right). \tag{7.7}$$

On the other hand, we observe that from the CLT for MLE estimations that for each *fixed* positive value v :

$$\overline{W}(v) \ge \lim_{n\to\infty} W_n(v) \ge \exp\left(-C\ v^2\right).$$

Example 7.3. Heavy tails of distributions.

We consider here the sample of a volume n from the standard one - dimensional Cauchy distribution: $X = R^1,\ \theta \in R^1, \theta_0 = 0,$

$$f(x,\theta) = \frac{\pi^{-1}}{1 + (x\ -\ \theta)^2}.$$

It is easy to calculate that

$$\phi(\lambda) \asymp C_1\ \lambda^2,\ |\lambda| \le C_2;$$

$$\phi(\lambda) \asymp C_3\ |\lambda|,\ |\lambda| \ge C_2.$$

More fine considerations as in the theorem 6.1 based on the exponential and power bounds for random fields maximum distribution based on the monograph [13], chapter 3, see also [14] show us that

$$W_n(v) \le \exp\left(-C_4\ v^2\right),\ v \le C_5;$$

$$W_n(v) \le C_6/v,\ v \ge C_5.$$

Therefore,

$$\sup_n W_n(v) \le C_7/v,\ v \ge C_8. \tag{7.8}$$

On the other hand,

$$\sup_n W_n(v) \ge W_1(v) = \mathbf{P}(|\xi(1)| > v) \ge C_9/v,\ v \ge C_{10}, \tag{7.9}$$

which coincides with upper bound (7.8) up to multiplicative constant.

At the same result is true for the symmetric stable distributions with the shift parameter $\theta: \ f(x,\theta) = f(x-\theta), \ \theta \in R^1$. In detail, let $\{f(\cdot,\cdot)\}$ be again the one - dimensional shift family of densities with characteristical functions

$$\int_{-\infty}^{\infty} e^{itx} \ f(x,\theta) \ dx = e^{it\theta \ - \ |t|^\alpha}, \ \alpha \in (1,2).$$

Using at the same arguments we obtain the following bilateral inequality:

$$C_1(\alpha)/v^\alpha \le \overline{W}(v) \le C_2(\alpha)/v^\alpha, \ v \ge 1.$$

Example 7.4. Scale parameter.

Let here $\{\xi(i)\}, \ i = 1,2,\ldots,n$ be a sample from the one - dimensional distribution $N(0,\theta), \ \theta > 0, X = R^1, \theta_0 = 1$.

The theorem 6.1 gives us the following estimation:

$$\overline{W}(v) \le \exp\left(-C \ v\right), v > 1. \tag{7.10}$$

The MLE $\hat{\theta}_n$ has an explicit view:

$$\hat{\theta}_n = n^{-1} \sum_{i=1}^n [\xi(i)]^2.$$

The distribution of $\hat{\theta}$ coincides, up to multiplicative constant, with the known χ^2 distribution with n degree of freedom.

We can see by means of this consideration that

$$\overline{W}(v) \ge \exp\left(-C \ v\right), v > 1; \tag{7.11}$$

and moreover for all values n

$$W_n(v) \ge C_1(n) \exp\left(-C_2(n) \ v\right), \ v \ge C_3(n). \tag{7.12}$$

At the same result is true for exponential distribution, indeed, when

$$f(x,\theta) = \theta^{-1} \ \exp\left(-x/\theta\right);$$

$X = R^1_+, \ \theta > 0, \ \theta_0 = 1$.

Notice that in this case the value λ_0 from the definition (2.6) is finite.

Remark 7.1

Note that the case of the so - called *penalized* modification of the MLE estimation (PMLE) may be considered analogously. See for definition and first results in the nonasymptotic risk estimations in the PMLE ([17]) and reference therein.

References

[1] Bagdasarova I.R. and Ostrovsky E.I. (1995). A nonuniform exponential estimators for large deviations in a Banach space. *Theory Probab. Appl.* **45** 638-642.

[2] Bednorz W. (2006). A theorem on Majorizing Measures. *Ann. Probab.* **34** 1771-1781. MR1825156

[3] Birge L., and Massart P. (1998). *Minimum contrast estimator on sieves: Exponential bounds and rates of convergence. Bernoulli*, 4(3): 113 - 150.

[4] Chernoff H. (1952). *A measure of efficiency for tests of a hypothesis Based on the sum of observations.* Ann. Math. Stat., 23: 493 - 507.

[5] Fernique X. (1975). Regularite des trajectoires des function aleatiores gaussiennes. *Ecole de Probablite de Saint-Flour, IV 1974, Lecture Notes in Mathematic.* **480** 1 96, Springer Verlag, Berlin.

[6] Golubev Yuri, Spokoiny Vladimir. *Exponential bounds for minimum contrast estimators.* Electronic publication, arXiv:0901.065v1 [math.ST] 6 Jan 2009.

[7] Ibragimov I.A.,and Khasminskij R.Z. (1981) *Statistical estimation. Asymptotic theory* New York - Heidelberg - Berlin; Springer Verlag.

[8] Jensen J.L. and Wood A.T. (1998) *Large deviations and other results for minimum contrast estimators.* Ann. Inst. Stat.Math., 50(4):673 - 695.

[9] Kozachenko Yu. V., Ostrovsky E.I. (1985). The Banach Spaces of random Variables of subgaussian type. *Theory of Probab. and Math. Stat.* (in Russian). Kiev, KSU, **32**, 43 - 57.

[10] Kurbanmuradov O., Sabelfeld K. (2007). Exponential bounds for the probability deviation of sums of random fields. Preprint. Weierstraβ - Institut fur Angewandte Analysis und Stochastik (WIAS), ISSN 0946 8633, p. 1-16.

[11] Ledoux M., Talagrand M. (1991) Probability in Banach Spaces. Springer, Berlin, MR 1102015.

[12] Ostrovsky E. Bide-side exponential and moment inequalities for tail of distribution of Polynomial Martingales. Electronic publication, arXiv: math.PR/0406532 v.1 Jun. 2004.

[13] Ostrovsky E.I. (1999). Exponential estimations for Random Fields and its applications (in Russian). Russia, OINPE.

[14] Ostrovsky E.I. (2002). Exact exponential estimations for random field maximum distribution. *Theory Probab. Appl.* **45** v.3, 281 - 286.

[15] Ostrovsky E., Rogover E. *Exact Exponential Bounds for the random Field maximum distribution via the Majorizing Measures (Generic chaining) method.* Electronic publication, arXiv:0802.0349v1 [math.PR] 4 Feb 2008.

[16] Seneta E. *Regularly Varying Functions*. Mir, Moscow edition, (1985).

[17] Spokoiny V. A penalized exponential risk bound in parametric estimation. Electronic publication, arXiv:0903.1721v1 [math.ST] 10 Mar 2009.

[18] Talagrand M. (1996). Majorizing measure: The generic chaining. *Ann. Probab.* **24** 1049 - 1103. MR1825156

[19] Talagrand M. (2001). Majorizing Measures without Measures. *Ann. Probab.* 29, 411-417. MR1825156

[20] Talagrand M. (2005). *The Generic Chaining. Upper and Lower Bounds of Stochastic Processes.* Springer, Berlin. MR2133757.

[21] Talagrand M.(1990). Sample boundedness of stochastic processes under increment conditions.*Ann. Probab.* **18**, 1 - 49.

[22] Van der Geer, S. (1993). *Hellinger - consistency of certain Nonparametric maximum likelihood estimators.* Ann. Stat., 21(1):14 - 44.